

## Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants

Article (Accepted Version)

Kelly, Taylor, Reby, David, Levréro, Florence, Keenan, Sumir, Gustafsson, Erik, Koutseff, Alexis and Mathevon, Nicolas (2017) Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants. *Biological Journal of the Linnean Society*, 120 (4). pp. 919-930. ISSN 0024-4066

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/63563/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Adult human perception of distress in the cries of bonobo,  
chimpanzee and human infants**

TAYLOR KELLY<sup>1</sup>, DAVID REBY<sup>2</sup>, FLORENCE LEVRÉRO<sup>3</sup>, SUMIR KEENAN<sup>3</sup>, ERIK  
GUSTAFSSON<sup>3,4</sup>, ALEXIS KOUTSEFF<sup>3,5</sup>, NICOLAS MATHEVON<sup>1,3\*</sup>

<sup>1</sup>*City University of New York, Hunter College, Department of Psychology, New York, USA*

<sup>2</sup>*University of Sussex, School of Psychology, Falmer, United Kingdom*

<sup>3</sup>*University of Lyon/Saint-Etienne, Equipe Neuro-Ethologie Sensorielle, ENES/Neuro-PSI  
CNRS UMR9197, Saint-Etienne, France*

<sup>4</sup>*University of Portsmouth, Department of Psychology, United Kingdom*

<sup>5</sup>*Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences - Center of  
PhenoGenomics, Lausanne, Switzerland*

\*Corresponding author. E-mail: mathevon@univ-st-etienne.fr

Running title: Perception of emotion in ape calls

## ABSTRACT

Understanding the extent to which humans perceive the emotional state of animals has both theoretical and practical implications. While recent studies indicate that natural selection has led to some convergence of emotion coding among vertebrate species (including humans), highlighting the interspecific value of emotional signals, it has also been argued that interspecific communication of emotions can fail due to species-specific signaling traits impairing information decoding, and/or absence of familiarity with heterospecific communication systems. Here we show that human listeners pay attention to the mean pitch of vocalizations when asked to rate the distress level expressed by human baby cries, and that they use a similar pitch scale to rate the emotional level of baby non-human ape (bonobo and chimpanzee) distress calls. As a consequence the very high-pitched bonobo infant calls were systematically rated as expressing overall high distress levels despite being recorded in contexts eliciting various stress intensity. Conversely, chimpanzee infant calls -which are characterized by a relatively lower-pitch- were systematically rated as expressing relatively lower distress levels. These results indicate that, in the absence of exposure/familiarity, our spontaneous ability to range the emotional content of vocalizations in closely related ape species remains biased by basic frequency differences, suggesting that the absolute interspecific value of emotional signals should not be over-estimated.

**KEYWORDS:** acoustic communication – evolutive convergence - ape – bonobo – chimpanzee - cry – distress call – emotion – human baby - signal.

## INTRODUCTION

By allowing to sense others' needs and respond in an appropriate manner, emotional recognition is a social facilitator in both humans and animals (Laukka & Juslin, 2005; Packard & Delafield-Butt, 2014). Deciphering how animals communicate their emotions is thus critical for understanding their social interactions. Since Darwin's book "The Expression of the Emotions in Man and Animal" (Darwin, 1872), many studies have explored the assumption that natural selection has led to convergences of emotion coding among animal species, conferring an interspecific value to emotional signals (Morton 1977; Panksepp 2005; Lingle & Riede, 2014). Much work has focused on the recognition of emotional valence (negative *versus* positive emotions) across species (Laukka *et al.*, 2005; Mendl *et al.*, 2010; Briefer, 2012). Yet emotions are not encoded discreetly and signals are usually graded to express various intensity levels or *arousal* (Laukka *et al.*, 2005; Briefer, 2012; Briefer *et al.*, 2015; Out *et al.*, 2010). Whether a universal emotional code allows fine assessment of the intensity level of an emotion across different species deserves further investigation. In the present study, we compare how human adult listeners rate the level of emotional distress expressed by human baby cries compared to that expressed by chimpanzee and bonobo baby cries. We hypothesise that although the general rule for coding distress may be shared among these three closely related primate species, humans' perception should primarily be tuned to the signals of their own offsprings. The species-specific features of vocalizations may thus impair the accurate assessment of heterospecific emotional signals in unfamiliar subjects.

An emotion is a short and intense reaction to stimuli, either of internal or external origin (Russell 2003; Laukka *et al.*, 2005). It is typically characterised by two dimensions: valence and arousal (LeDoux 1995; Mendl *et al.*, 2010; Briefer 2012; but see Laukka *et al.*, 2005 for additional dimensions). Emotional valence is the intrinsic attractiveness (positive valence) or aversiveness (negative valence) of a situation leading to an emotion. Emotional arousal is the level of alertness, excitement, or engagement produced by the objects of emotion (Goudbeek & Scherer, 2010; Briefer *et al.*, 2014). Because emotions affect the autonomic nervous system and impact the muscles used to produce vocalizations, vocal communication is a privileged channel for communicating emotions and there is evidence for acoustic correlates of both emotional dimensions (Johnstone & Scherer 2000). Negative valence is usually characterized by vocalizations with a higher amplitude, with more energy in the higher part of the frequency spectrum, a wider frequency range and more spectral noise (Fichtel & Hammerschmidt, 2002; Waaramaa *et al.*, 2010; Hammerschmidt & Jurgens, 2007;

Goudbeek & Scherer, 2010). Increase in arousal level is associated with longer, louder and harsher vocalizations with more energy in the higher part of frequency spectrum and a more pronounced frequency modulation (Blumstein & Recapet, 2009; Fichtel & Hammerschmidt, 2002). These acoustic correlates follow Morton's "motivation-structural" rules hypothesis (Morton, 1977) stating that the physical structure of vocalizations converged between species in response to identical natural selection constraints.

Distress is a negative emotion of variable arousal, from anxiety to panic fear (Goudbeek & Scherer, 2010), and its expression can have a high survival value, e.g. in the context of parent-offspring communication. The expression of distress in infant vocalizations has retained the attention of several studies (Wood & Gustafson, 2001; Lingle *et al.*, 2012; Newman, 2007), and it has been demonstrated that distress calls (e.g. the "cries" of human babies) share similar acoustical features across species and elicit functional responses from caregivers (Lingle & Riede, 2014; Newman, 2007). This shared acoustic coding of distress explains the interspecific response that has been observed in some animals and human listeners when testing individuals with distress vocalizations from infants of another species (e.g. in deer, Lingle & Riede, 2014). It is however established that such interspecific communication has limitations as species-specific traits can impair listeners' sensitivity to the distress information embedded in the vocalizations. For instance, female deer (*Odocoileus hemionus* and *Odocoileus virginianus*) do not respond to newborn distress calls of other species if the fundamental frequency (the "pitch") is outside the deer's "frequency response range", i.e., approximately  $\pm 50\%$  of the mean pitch of a conspecific infant's distress call (Lingle & Riede, 2014; Teichroeb *et al.*, 2013). Besides it has been argued that interspecific emotional recognition in humans could be shaped by experience (familiarity with the animal species) more than by universal acoustic coding (Scheumann *et al.*, 2014).

Distress vocalizations in animals and human cries are usually graded according to the level of emotional arousal (Lingle *et al.*, 2012; Briefer, 2012, Out *et al.*, 2010), giving caregivers information about the level of urgency experienced by the baby, and thus of primary survival value (Briefer, 2012; Out *et al.*, 2010; Gustafsson *et al.*, 2013; Lindova *et al.*, 2015). It is well established that human caregivers demonstrate fine perception of information coded in babies' cries and assess urgency using cry's dynamic cues by reacting more promptly to louder, higher-pitched and less voiced cries (Briefer, 2012; Out *et al.*, 2010; Gustafsson *et al.*, 2013; Lindova *et al.*, 2015; Reby *et al.*, 2016). Despite the importance of graded coding in communicating distress levels, most previous studies investigating

interspecific response to distress calls ignore this aspect and focused on the ability to categorize distress vocalizations against other emotional signals (Lindova *et al.*, 2015).

Chimpanzee *Pan troglodytes* and bonobo *Pan paniscus* are the closest living relatives to humans (Prüfer *et al.*, 2012), and mothers of both ape species offspring respond to their offspring's vocalizations by bringing assistance (Goodall, 1986; Bermejo & Omedes, 1999). The auditory channel plays an important role in mother-offspring communication in forest habitats, especially when infants gain in independence and make more exploratory trips away from their mother (Goodall, 1986). As human babies' cries, ape infant distress calls contain dynamic cues which vary with the level of distress, as well as static cues to individual identity that may allow recognition by the mother (Levréro & Mathevon, 2013). While human and non-human ape infant crying clearly share many similarities, to our knowledge the acoustic structure of distress vocalizations of human and non-human ape infants has never been directly compared, and how human listeners perceive distress in the infants' calls of their closest relatives remains to be investigated. The present study thus represents an opportunity to test our spontaneous ability to perceive the emotional content of other species' vocalizations.

Here we investigated how human adult listeners rate the level of distress expressed in infants' distress calls of human and non-human ape (bonobo and chimpanzee). We compared the acoustic structure of distress vocalisations, and performed psychoacoustic tests on human adult listeners. We predicted that interspecific pitch differences (bonobo calls are noticeably higher-pitched than chimpanzee and human vocalizations (de Waal, 1988) would bias human listeners.

## METHODS

### Sound signals

Human babies' cries (Fig.1) were collected in two different conditions: bathing (low arousal context, Supplementary Recording S1) and vaccination (high arousal context, Supplementary Recording S2). Babies (N = 24 individuals with 12 recorded in the "bath" context and 12 in the "vaccination" context, balanced sex ratio, 2-5 months old) were recorded with a Sennheiser MD42 microphone, placed at 30 cm from the baby and connected to a digital

recorder (Zoom H4n, sampling frequency = 44100 Hz). “Bath” cries were obtained while babies were given their bath by their parents at home. “Vaccine” cries were recorded during scheduled routine vaccination at the pediatrician’s office (vaccine = Prevnar©, Pneumococcal 7-valent conjugate; Pfizer Ltd, Sandwich, Kent, United Kingdom). For the purpose of analysis and playback experiments, we isolated crying sequences lasting 6-8 seconds for each of the 24 recordings.

Bonobo infant distress calls (Fig.1, Supplementary Recording S1) were recorded with a Sennheiser MKH70-1 ultra-directional microphone connected to a Zoom H4n recorder (sampling frequency = 44100 Hz). Recordings were performed in three zoological parks (Apenheul in Apeldoorn, Netherlands; Planckendael in Mechelen, Belgium; la Vallée des Singes in Romagne, France) from 2011 to 2014. Recorded infants (N = 6 individuals aged from 1 to 4 year-old) were all dependent of their mother (fully or part time breast fed and frequently carried by the mother). From the recordings, we isolated 12 sequences lasting 6-8 seconds (1-8 sequences/individual) for the acoustic analysis and for the psycho-acoustic experiments.

Chimpanzee infant distress calls (Fig. 1, Supplementary Recording S4) were collected with a Sennheiser MKH70-1 ultra-directional microphone connected to a Marantz PMD670 digital recorder (sampling frequency = 44100 Hz) in 2009. Recorded infants (N = 7 individuals) come from a wild population (Kanyanchu community in Kibale National park, Uganda) habituated to humans since 1991 for ecotourism (Johns, 1996). Because they are wild animals, their exact age is not known but all were under 4 year-old and have been seen carried by their mother. From the recordings, we isolated 12 sequences lasting 6-8 seconds (1-3 sequences/individual).

For both apes, the contexts of recordings went from begging the mother for assistance in the absence of any visible danger (N = 5 recordings for bonobos, N = 5 recordings for chimpanzees) to calling for being picked up when infants were involved in aggressive interactions (N = 5 for bonobos, N = 2 for chimpanzees). The context of remaining recordings was unclear but always characterized by an infant soliciting the mother. For practical and ethical reasons, it was not possible to quantify objectively the emotional state of ape babies through physiological measurements (e.g. heart rate, Briefer et al. 2015). It was thus not possible to reliably and objectively estimate the arousal intensity expressed by their vocalizations. For the purpose of the present study, we thus chose to not attribute any *a priori* arousal value to any of the recordings. However, given the conditions in which recordings

took place in captive or habituated populations allowing intense monitoring and a variety of recording contexts, we assume that our recordings represent a fair distribution of arousal levels coded by ape babies' distress calls.

### **Acoustic analysis**

We characterized the acoustic structure of the 48 selected sequences using a set of 15 temporal and spectral parameters. We performed acoustic analyses using a dedicated batch-processing script in PRAAT (Boersma & Weenink, 2015), which contained four distinct procedures.

The first procedure of the script characterized the pitch and its intonation (pitch contour variation) of the vocalizations. The pitch contour was extracted using the To Pitch(cc) command, and the following parameters were derived: %voiced (percentage of the signal that is characterized by a detectable pitch), mean pitch, max pitch, min pitch (respectively the mean, maximum and minimum pitch calculated over the duration of the signal) and pitchCV (coefficient of variation of the pitch over the duration of the signal). In a second step, two distinct smoothing algorithms (Smooth command in Praat) were performed on the pitch contour: the first allowed a relatively broad bandwidth (Smooth command parameter), to suppress very short-term frequency fluctuation while preserving minor intonation events (such as bleat-like frequency modulation), and the second only allowed a narrow bandwidth (Smooth command parameter), to only characterize strong pitch modulation (major intonation events). Inflection points were counted (as each change in the sign of the contour's derivative) after each smoothing procedure, and divided by the total duration of the voiced segments in each recording, resulting in two distinct indexes of pitch variation (inflex25 and inflex2). A second procedure focused on the intensity contour and allowed the characterization of the variability of the cries' intensity by calculating intCV, the coefficient of variation of the intensity contour estimated using the To intensity y command in PRAAT. A third procedure focused on the periodic quality of the signal and measured the harmonicity (harm, degree of acoustic periodicity, measured as the ratio of harmonics to noise in the signal and expressed in dB), an index of jitter (jitter, small fluctuation in periodicity measured as the average of 'local', 'rap' and 'ppq5' measures in PRAAT) and an index of shimmer (shimmer, small variation in amplitude between consecutive periods, measured as the average of 'local', 'apq5' and 'apq11' parameters in PRAAT). A final procedure characterized the spectral envelope of the cry by applying a cepstral smoothing procedure (bandwidth: 900 Hz) to each



recorded sequence, followed by the extraction of the first four spectral prominences (SP1, SP2, SP3, SP4) of the resulting smoothed spectrum. Because human –and possibly bonobo and chimpanzee- vocalizations can be strongly nasalized, and/or can contain biphonation phenomena (Soltis, 2004; Riede *et al.*, 2004) that can create resonance-independent broadband components, the measured spectral peaks cannot be safely considered as accurate measure of formant frequencies and are therefore termed spectral prominences.

To illustrate differences between categories of emitters (human –bath and vaccine contexts-, bonobo and chimpanzee), we first performed a multivariate principal component analysis (PCA) to reduce the 15 non-independent acoustic parameters described above into two independent Acoustic Dimension (AD1 and AD2). We further used linear regression models (LM; package lme4, R version 3.1.2) with AD1 and AD2 as dependent measures (fixed effect: category of emitters). Post-hoc multiple comparison tests (Tukey contrasts, R package multcomp) were used to compare between categories of emitters.

## **Psychoacoustic experiments on human listeners**

Twenty-one women and nine men ( $30 \pm 14$  years old, min = 18, max = 64), recruited through personal requests, took part in the study. Each of these participants had no previous experience with bonobos or chimpanzees, but all had previous experience with human infants.

Each participant was tested alone in a quiet room. After being given brief instructions, he/she listened to the sound stimuli through headphones (Sennheiser HD 201 Closed Back Headphones) connected to a laptop computer using the Experiment Multiple Forced Choice Tool in PRAAT software, and was asked to rate a succession of sound stimuli (human baby, bonobo infant and chimpanzee infant cries/calls). Participants entered each rating by clicking on the chosen button on the screen, then they could either confirm their choice (“OK” button), replay the sound (“REPLAY” button) or change their rating (“OOPS” button). The playback test was conducted as a double-blind experiment.

Participants were asked to rate the distress expressed by the stimuli according its level of arousal. As distress varies in arousal from anxiety to panic fear (Goudbeek & Scherer (2010)), the rating scale proposed to the participants was a Likert scale ranging from 1 to 7, with: 1 = no anxiety, 2-3 = anxiety, 4 = medium fear, 5-6 = strong fear, 7 = panic/fear. Participants were tested with the entire collection of stimuli (12 human baby “bath” cries, 12 human baby “vaccine” cries, 12 bonobo infant calls, 12 chimpanzee infant calls; duration of

6-8 seconds each; normalized intensity). The stimuli presentation was randomized. Participants were blind as to whether they were listening to a human, a chimpanzee, or a bonobo infant vocalization.

## **Ethical statement**

Research methods were approved by the ethic committee of the CHU of Saint-Etienne (IRBN672015/CHUSTE) and the CNIL committee of the university of Saint-Etienne. Informed consent was obtained from all adult subjects.

## **Statistical analysis of psychoacoustic experiments**

To test for the influence of the vocalizations' acoustic features on human listeners' ratings, we used linear mixed effect models with rating level by listeners as a dependent measure. We built two separate models. In the first one, the acoustic variables were the vocalizations' two first dimensions defined by the Principal Component Analysis described above (fixed effects: acoustic dimensions, listener's gender; random effects: listener identity, acoustic stimulus; package lme4, R version 3.1.2). The listener's gender was included as a fixed effect because some previous studies suggested an effect of gender on the assessment of infant cries by adult listeners (e.g. De Pisapia *et al.* 2013, but see Gustafsson *et al.* 2013). As reported in Results below, neither the listener's gender nor the interactions between gender and acoustic dimensions had a significant effect on listeners' ratings in the present study. We therefore pooled male and female listeners in the subsequent analysis. In the second model, the acoustic variables were the 15 temporal and spectral features (fixed effect: acoustic variable; random effects: listener identity, acoustic stimulus). All models had a maximal random effects structure, with random slopes within the listener's identity for each acoustic parameter (Barr *et al.*, 2013). P values were obtained with likelihood-ratio tests comparing the fit of full models with reduced models lacking the fixed effect.

To test for the influence of the origin of recordings (i.e., bonobo, chimpanzee, human "bath", human "vaccine") on listeners' ratings, we used a linear mixed effect model with rating level as a dependent measure (fixed effects: recorded condition; random effects: listener identity, acoustic stimulus). The recorded condition was included as a random slope

within the listener's identity. P values were obtained with likelihood-ratio tests comparing the fit of full models with reduced models lacking the fixed effect. This analysis was followed by post-hoc multiple comparison tests (function glht in multcomp R package).

## RESULTS

### Acoustic characteristics of vocalizations

The principal component analysis performed on the 15 acoustic variables characterizing the acoustic structure of calls highlighted significant differences between species and recording conditions (Fig.2, N = 12 calls for each four categories; LM on AD1:  $F(3, 44) = 39.2$ ,  $p < 0.001$ ; LM on AD2:  $F(3, 44) = 10.4$ ,  $p < 0.001$ ; multiple comparisons for AD1 and AD2: see Table 1). Fig.2 illustrates that human cries recorded in both the bath and vaccine contexts show closer acoustic features, whereas chimpanzees' and bonobos' calls are more similar to each other.

As underlined in Table 2, the first dimension of the acoustic space (AD1) defined by the principal component analysis is mainly related to pitch and periodic quality of signals. Bonobos' calls are the most high-pitched vocalizations. The pitch of chimpanzees' calls are more similar to human babies' cries in the vaccination context, while human babies in the bath context are the most low-pitched vocalizations (Table 3). Temporal fluctuations in periodicity (jitter) and amplitude (shimmer) follow the same pattern. The second dimension (AD2) is mostly explained by the spectral prominences: human cries recorded in the vaccine context and bonobos' calls are both characterized by a reinforced energy amount in the higher part of their frequency spectrum (Table 3).

### Assessment of calls' emotion content by human listeners

When rating vocalizations, listeners relied mainly on the first dimension defining the acoustic space of vocalizations (AD1, see above), while neither the second dimension (AD2) nor the interaction between AD1 and AD2 were significant predictors of the rating (Table 4). In accordance with this result we found that mean pitch, the most prominent acoustic parameter weighting on AD1 (Table 2), represents also the most prominent predictor of listeners' rating of fear (Table 4). Neither listener's gender nor the interactions between listener's gender and AD1 or AD2 were significant predictors of the rating (Table 4).

The nature of recordings - human, bonobo or chimpanzee vocalizations - significantly influenced the participants' rating (GLM:  $\chi^2 = 42.5$ ,  $df = 3$ ,  $P < 0.001$ ). Specifically, bonobo calls were rated as expressing higher distress than chimpanzee calls (rating of bonobo calls: mean  $\pm$  SD =  $5.32 \pm 0.94$ ; rating of chimpanzee calls =  $3.79 \pm 1.16$ ; multiple comparisons test:  $Z = -5.00$ ,  $P < 0.001$ ) and than human baby cries (rating of "bath" cries =  $2.92 \pm 0.79$ ; comparison with bonobo calls:  $Z = -7.63$ ,  $P < 0.001$ ; rating of "vaccine" cries =  $4.06 \pm 0.99$ ; comparison with bonobo calls:  $Z = -3.91$ ,  $P < 0.001$ ). Human baby "vaccine" cries were also rated as expressing higher distress than human baby "bath" cries ( $Z = 3.58$ ,  $P = 0.0019$ ). Conversely, the rating of human baby "vaccine" cries was not significantly different from the chimpanzee calls' rating ( $Z = 0.856$ ,  $P = 0.827$ ), while the rating of human baby "bath" cries and chimpanzee calls were significantly different ( $Z = -2.63$ ,  $P = 0.043$ ).

As a result, human listeners rated the distress level of human babies' cries according to their mean pitch with the highest pitch rated as the more distressful, and apply the same scaling rule to ape distress calls (Fig. 3).

## DISCUSSION

The main aim of this study was to contrast how human listeners rate human's and apes' infant distress vocalizations. We showed that calls of the three species differ on the basis of their acoustic structure. Psycho-acoustic experiments demonstrated that human listeners rely mainly on pitch to assess the level of distress, with higher pitched vocalizations rated as more distressful. Using this rule, human listeners were able to range human babies' cries from low to high arousal, while bonobo calls were all classified as expressing high levels of distress. Yet, bonobo calls had been recorded in different contexts eliciting various levels of stress. Chimpanzee calls - which are lower pitched than bonobos' - were classified as expressing less distress although they were recorded in contexts associated with similar stress levels. Although the absence of objective assessment of the distress level experienced by the primate infants during our recordings (discussed further below) limits our ability to attribute acoustic variability and listeners' ratings to specific distress levels, the results reveal that the ability of non-expert human listeners to spontaneously range the emotional content of apes' vocalizations may be rather limited.

### Infant vocalizations differ between species

The acoustical analysis of vocalizations showed that human and ape infants' calls significantly differ by their acoustic features, with both species and level of stress as modulating factors. Adult bonobos are well-known for their high pitched vocalizations (de Waal, 1988), and our analysis shows that bonobo infant calls are in line with adult calls: the recordings we used were characterized by a very high pitch, mostly out of the range of the human babies' cries. The pitch of chimpanzees' recordings was lower, and closer to the one of human babies. However and quite interestingly, non-human apes' distress vocalizations were more similar to each other than to the human babies' cries within the 2D acoustic space defined by the two first principal components: this acoustic proximity may reflect the anatomical proximity of the vocal apparatus between this two very closely related species (Lieberman *et al.*, 1972). A previous study has demonstrated that phylogenetic trees reconstructed from the acoustic features of vocalizations associated with laughter –another emotional signal expressing a positive emotional valence- do match the trees based on genetics (Ross *et al.*, 2009). Although we did not conduct such an analysis with our recordings, the acoustic proximity we observed between bonobos and chimpanzees (Fig.2) supports the possibility that infant distress calls may also express a phylogenetic continuity within apes and with the human species.

Previous studies of mammal distress vocalizations have emphasized a general trend for distress coding that is characterized by sound pitch increasing with arousal (Briefer, 2012; Zimmermann *et al.*, 2013; Linhart *et al.*, 2015). As underlined by Maruscakova *et al.* (2015), this general rule supports theories of shared emotional systems across mammalian species (Darwin, 1872; Morton, 1977). Assessing the pitch of a sound is a relatively straightforward process that allows categorizing signals according both their valence (positive versus negative) and intensity or arousal (Briefer, 2012). Human listeners thus assess the emotional intensity of both human verbal and non-verbal sound signals using their pitch (Porter *et al.*, 1986; Laukka *et al.*, 2005; Sauter *et al.*, 2010), and use this “pitch-rule” to rate the emotional content of vocalizations from other mammal species with reasonable accuracy (e.g. piglet's calls, Maruscakova *et al.*, 2015; dogs' calls, Farago *et al.*, 2014; cats' solicitation purrs, McComb *et al.*, 2009). In the present study, ratings of human babies' calls support the importance of pitch for attribution of emotional arousal in human listeners: cries with higher arousal (recorded during vaccination) were higher pitched and indeed rated as expressing more distress than cries with lower arousal (recorded at bath).

**The “pitch-rule” followed by human listeners is likely to impair interspecific assessment of call emotional content**

Following this “pitch rule” (Maruscakova *et al.*, 2015), human listeners rated bonobo and chimpanzee infant calls accordingly to their mean pitch level. As a consequence bonobo calls were rated as expressing very high level of distress compared to both human baby and chimpanzee calls. Even though it was not possible to objectively assess the actual distress experienced by these infants in the present study (preventing us from investigating potential correlations between the pitch of bonobo calls and distress), this homogeneous rating by human listeners is unlikely to reflect an actual interspecific difference in the level of distress experienced by infants. Indeed, bonobo infants were recorded in contexts likely to elicit a range of stress levels, and chimpanzee infants’ vocalizations recorded in similar contexts were not rated as expressing as much distress as bonobos’ calls. Finally, human listeners classified chimpanzee calls at the same level as human babies’ vaccine-elicited cries.

Previous studies have indicated that human listeners could make errors in assessing the emotional content of other species’ vocalizations. For instance, human listeners attribute an emotional valence to tree shrew calls that is opposite to the actual valence (Scheumann *et al.*, 2014). This result was attributed to the lack of familiarity of humans with this mammal species. This error could also be linked to the fact that the acoustic characteristics –and specifically the pitch- of the tree shrew vocalizations are well out of the range of human vocalizations. In their study, Lingle & Riede (2014) showed that female deer only recognize distress coded in heterospecific vocalizations if their mean pitch is manipulated to fall within the frequency range characterizing their own offspring’ distress calls. We predict that shifting the frequency of bonobo calls downwards would lead to more realistic assessments by human listeners.

Finally, our analysis suggests that the relationship between emotions and their expression through vocalizations does not solely involve pitch levels and that other parameters related to sound periodicity (harm, jitter, shimmer) and dynamics of amplitude modulation (intCV) are likely to co-vary with distress levels. Bonobo and chimpanzee adults may pay attention to some of these other acoustic features to decode the distress level embedded in their infants’ calls. In a recent study of piglet calls, Linhart *et al.* (2015) emphasized that different acoustic features encode arousal in two different call types (the scream and the grunt). While the pitch appears as a generally widely shared parameter used by listeners to decode distress, its actual reliability may be limited and other parameters may

play more significant roles in encoding distress in across vertebrate vocal systems (Arnal *et al.*, 2015).

### **Limitations and perspectives**

A limitation of the present study is the absence of objective assessment of the distress level experienced by the bonobo and chimpanzee infants during the recording of their vocalizations. To establish if these two species encode distress according to a pitch scale rule, it would be necessary to perform a physiological assessment of stress, such as measure of heart rate or assessment of cortisol levels. Moreover, we cannot ascertain that the recordings from chimpanzees and bonobos cover the same range of distress levels with similar weightings. However, our recordings represent a range of contexts likely to represent a diversity of distress levels in both non-human primates. Although we did not measure physiological indicators in human babies, we believe that it is reasonable to consider that the two conditions (bathing and vaccination) elicit categorically different levels of distress and arousal, as one condition is mildly unpleasant and common, while the other is both painful and uncommon.

Our results indicate that human listeners attributed more variable ratings (spanning more widely across the scale - Fig.3) to human baby calls. While this may indicate that the distress levels experienced by the human babies were more varied than those experienced by the non-human primates during our recordings, this may also reflect limitation in human listeners' ability to perceive and interpret variability in unfamiliar heterospecific signals.

An interesting area for further studies would be to investigate the effect of familiarity on human listeners' ability to rate bonobo and chimpanzee calls. In the present study, human subjects were familiar to human babies but had no previous experience of either bonobos or chimpanzees. It is well established that experience can affect the perception of human babies' cries (e.g. caregivers spending more time with their baby become more efficient in recognizing her/him from her/his cries alone, Gustafsson *et al.*, 2013). It is therefore possible that familiarity with non-human ape vocalisations may increase human listeners' ability to perceive subtle acoustic variation linked to distress intensity, thereby overcoming basic biases linked to interspecific variation in pitch. Experiments with experienced listeners (e.g. zookeepers) would be necessary to test the effect of experience when assessing stress in the acoustic signals of non-human species.

Further studies could also investigate the multimodal dimensions of infant distress communication. Indeed, other cues such as facial expressions or body movements may contribute to the assessment of distress.

## ACKNOWLEDGEMENTS

We thank the Uganda Wildlife Authority and the Uganda National Council for Science and Technology for permission for allowing us to conduct the recordings of chimpanzees in the Kibale National Park. We are grateful to J. Kasenene, director of Makerere University Biological Field Station, S. Kyamukama N. Magambo and R. Mchwezi, assistant researchers, for their help in data collection in the field, and S. Krief for her help to access to the field. We warmly thank L. Déruti, J. Ruijs, J. Stevens, J.P Guéry and the keepers of Apenheul, Planckendael and la Vallée des Singes zoos for their assistance and support, Dr.Olivier Martin for help during the recordings of human babies, the parents who kindly accepted to participate in this study and their babies, the adults who participated in the psycho-acoustic experiments, and Nicolas Boyer for his technical assistance. We thank three anonymous reviewers for their helpful comments. This work was funded by the university of Lyon/Saint-Etienne. Co-author David Reby was invited professor at the University of Lyon/Saint-Etienne during part of this research project.

## AUTHOR CONTRIBUTIONS

Designed the study: TK, NM. Did the recordings: EG, SK, AK, FL. Did the psycho-acoustic experiments: TK. Analyzed the data: TK, NM, DR. Wrote the paper: TK, FL, NM, DR. All authors contributed to its revision.

## REFERENCES

**Arnal LH, Flinker A, Kleinschmidt A, Giraud AL, Poeppel D. 2015.** Human screams occupy a privileged niche in the communication soundscape. *Current Biology* **25**: 2051-2056.



453 **Barbosa MN, Mota MT. 2014.** Do newborn vocalizations affect the behavioral and  
 454 hormonal responses of non-reproductive male common marmosets (*Callithrix jacchus*)  
 455 *Primates* **55(2)**: 293-302.

456 **Barr DJ, Levy R, Scheepers C, Tily HJ. 2013.** Random effects structure for confirmatory  
 457 hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**: 255-278.

458 **Bermejo M, Omedes A. 1999.** Preliminary vocal repertoire and vocal communication of wild  
 459 bonobos (*Pan paniscus*) at Lilungu. *Folia Primatologica*. **70**: 328-357.

460 **Blumstein D, Récapet C. 2009.** The Sound of Arousal: The Addition of Novel Non-  
 461 linearities Increases Responsiveness in Marmot Alarm Calls. *Ethology*, **115(11)**: 1074-1081.

462 **Boersma P, Weenink D.** Praat: doing phonetics by computer (Version 6.0.04) [Computer  
 463 program]. Retrieved from <http://www.fon.hum.uva.nl/praat/> (2015).

464 **Briefer E. 2012.** Vocal expression of emotions in mammals: Mechanisms of production and  
 465 evidence. *Journal of Zoology* **288(1)**: 1-20.

466 **Briefer E, Maigrot AL, Mandel R, Briefer SF, Bachmann I, Hillmann E. 2015.**  
 467 Segregation of information about emotional arousal and valence in horse whinnies. *Scientific*  
 468 *Reports* **4**.

469 **Darwin C. 1872.** The expression of the emotions in man and animal. London, England:  
 470 Murray.

471 **De Pisapia N, Bornstein MH, Rigo P, Esposito G, De Falco S, Venuti P.** Gender  
 472 differences in directional brain responses to infant hunger cries. *Neuroreport*, **24**: 142-146.

473 **De Waal FBM. 1988.** The communicative repertoire of captive bonobos (*Pan paniscus*),  
 474 compared to that of chimpanzees. *Behaviour*, **106**:183-251.

475 **Fichtel C, Hammerschmidt K. 2002.** Responses of Redfronted Lemurs to Experimentally  
 476 Modified Alarm Calls: Evidence for Urgency- Based Changes in Call Structure. *Ethology*,  
 477 **108(9)**: 763-778.

478 **Fleming AS, Corter C, Stallings J, Steiner M. 2002.** Testosterone and Prolactin Are  
 479 Associated with Emotional Responses to Infant Cries in New Fathers. *Hormones and*  
 480 *Behavior* **42(4)**: 399-413.

481 **Goodall J. 1986.** *The chimpanzees of Gombe: patterns of behaviour*. Cambridge, MA:  
 482 Harvard University Press.

483 **Goudbeek M, Scherer K. 2010.** Beyond arousal: Valence and potency/control cues in the  
 484 vocal expression of emotion. *The Journal of the Acoustical Society of America*, **128(3)**: 1322-  
 485 36.

486 **Gustafsson E, Levréro F, Reby D, Mathevon N. 2013.** Fathers are just as good as mothers  
 487 at recognizing the cries of their baby. *Nature Communications* **4**: 1698.

488 **Hammerschmidt K, Jürgens U. 2007.** Acoustical correlates of affective prosody. *Journal of*  
 489 *Voice* **21**: 531-540.

490 **Johns BG. 1996.** Responses of chimpanzees to habituation and tourism in the Kibake Forest,  
 491 Uganda. *Biology Conservation* **78**: 257-262.

492 **Johnstone T, Scherer KR. 2000.** Vocal communication of emotion. In « The Handbook of  
 493 Emotion » (Eds. Lewis M & Haviland J). New York, Guilford.

494 **Laukka P, Juslin P, Bresin R. 2005.** A dimensional approach to vocal expression of  
 495 emotion. *Cognition & Emotion* **19(5)**: 633-653.

496 **Laukka P, Juslin PN. 2005.** A dimensional approach to vocal expression of emotion.  
 497 *Cognition and Emotion* **19**: 633-653.

498 **LeDoux JE. 1995.** Emotion : Clues from the brain. *Annual Review of Psychology* **46**:209-  
 499 235.

500 **Levréro F, Mathevon N. 2013.** Vocal Signature in Wild Infant Chimpanzees. *American*  
 501 *Journal of Primatology* **75(4)**: 324-332.

502 **Lieberman P, Crelin ES, Klatt DH. 1972.** Phonetic ability and related anatomy of the  
 503 newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*  
 504 **74**: 287-307.

505 **Lindova J, Špinka M, Nováková L. 2015.** Decoding of Baby Calls: Can Adult Humans  
 506 Identify the Eliciting Situation from Emotional Vocalizations of Preverbal Infants? *PLOS*  
 507 *ONE* **10(4)**.

508 **Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA. 2012.** What makes a cry a  
 509 cry? A review of infant distress vocalizations. *Current Zoology* **58(5)**: 698-726.

510 **Lingle S, Riede T. 2014.** Deer mothers are sensitive to infant distress vocalizations of diverse  
 511 mammalian species. *The American Naturalist* **184(4)**: 510-522.

512 **Linhart P, Ratcliffe VF, Reby D, Spinka M. 2015.** Expression of emotional arousal in two  
513 different piglet call types. *PloS ONE* **10**: e0135414.

514 **Marler P. 1957.** Species distinctiveness in the communication signals of birds. *Behaviour*  
515 **11**:13-39.

516 **Maruscakova IL, Linhart P, Ratcliffe VF, Tallet C, Spinka M. 2015.** Humans (Homo  
517 sapiens) judge the emotional content of piglet (*Sus scrofa domestica*) calls based on simple  
518 acoustic parameters, not personality, empathy, nor attitude toward animals. *Journal of*  
519 *Comparative Psychology* **129**: 121-131.

520 **Mendl M, Burman O, Paul E. 2010.** An integrative and functional framework for the study  
521 of animal emotion and mood. *Proceedings Biological Sciences of The Royal Society*  
522 **277(1696)**: 2895-904.

523 **Newman J. 2007.** Neural circuits underlying crying and cry responding in mammals.  
524 *Behavioural Brain Research*, **182(2)**: 155-165.

525 **Out D, Pieper S, Bakermans-Kranenburg MJ, Zeskind PS, Ijzendoorn MH. 2010.**  
526 Intended sensitive and harsh caregiving responses to infant crying: The role of cry pitch and  
527 perceived urgency in an adult twin sample. *Child Abuse & Neglect* **34(11)**: 863-873.

528 **Packard A, Delafield-Butt. 2014.** Feelings as agents of selection: putting Charles Darwin  
529 back into (extended neo-) Darwinism. *The Biological Journal of the Linnean Society* **112(2)**:  
530 332-353.

531 **Panksepp J. 2005.** Affective consciousness: Core emotional feelings in animals and humans.  
532 *Consciousness and Cognition* **14**: 30-80.

533 **Parr LA, Waller BM, Fugate J. 2005.** Emotional communication in primates: Implications  
534 for neurobiology. *Current Opinion in Neurobiology* **15(6)**: 716-720.

535 **Porter FL, Miller RH, Marshall RE. 1986.** Neonatal Pain Cries: Effect of Circumcision on  
536 Acoustic Features and Perceived Urgency. *Child Development* **57(3)**: 790–802.

537 **Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G,**  
538 **Kodira C, Winer R, Knight JR, Mullikin JC, Meader SJ, Ponting CP, Lunter G,**  
539 **Higashino S, Hobolth A, Dutheil J, Karakoç E, Alkan C, Sajjadian S, Catacchio CR,**  
540 **Ventura A, Marques-Bonet T, Eichler EE, André C, Atencia R, Mugisha L, Junhold J,**  
541 **Patterson N, Siebauer M, Good JM, Fischer A, Ptak SE, Lachmann M, Symer DE,**

542 **Mailund T, Schierup MH, Andr  s AM, Kelso J, P   bo S. 2012.** The bonobo genome  
543 compared with the chimpanzee and human genomes. *Nature* **486**: 527-531.

544 **Reby D, Levrero F, Gustafsson E, Mathevon N. 2016.** Sex stereotypes influence adults'  
545 perception of babies' cries. *BMC Psychology*, **4**:19.

546 **Riede T, Owren MJ, Arcadi AC. 2004.** Nonlinear acoustics in pant hoots of common  
547 chimpanzees (*Pan troglodytes*): frequency jumps, subharmonics, biphonation, and  
548 deterministic chaos. *American Journal of Primatology* **64(3)**: 277-291.

549 **Rilling JK. 2013.** The neural and hormonal bases of human parental care. *Neuropsychologia*  
550 **51(4)**: 731-747.

551 **Ross MD, Owren MJ, Zimmermann E. 2009.** Reconstructing the evolution of laughter in  
552 great apes and humans. *Current Biology* **19**: 1106-1111.

553 **Russell JA. 2003.** Core affect and the psychological construction of emotion. *Psychological*  
554 *Review* **110**: 145-172.

555 **S  nchez SM, Ziegler TE, Snowdon CT. 2014.** Both parents respond equally to infant cues  
556 in the cooperatively breeding common marmoset, *Callithrix jacchus*. *Animal Behaviour* **97**:  
557 95-103.

558 **Sauter DA, Eisner F, Ekman P, Scott SK. 2010.** Cross-cultural recognition of basic  
559 emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of*  
560 *Sciences of the United States of America* **107**: 2408-2412.

561 **Scheumann M, Hasting AS, Kotz SA, Zimmermann E. 2014.** The Voice of Emotion across  
562 Species: How Do Human Listeners Recognize Animals' Affective States? *PLoS ONE* **9(3)**.

563 **Soltis J. 2004.** The signal functions of early infant crying. *Behavioral and Brain Sciences* **27**:  
564 443-458.

565 **Sueur J, Aubin, T, Simonis C. 2008.** Equipement review: seawave, a free modular tool for  
566 sound analysis and synthesis. *Bioacoustics* **18**: 213-226.

567 **Teichroeb LJ, Riede T, Kotrba R, Lingle S. 2013.** Fundamental frequency is key to  
568 response of female deer to juvenile distress calls. *Behavioural Processes* **92**: 15-23.

569 **Waal FB. 1988.** The Communicative Repertoire of Captive Bonobos (*Pan Paniscus*),  
570 Compared To That of Chimpanzees. *Behaviour* **106(3)**: 183-251.

571 **Waaramaa T, Laukkanen AM, Airas M, Alku P. 2010.** Perception of emotional valences

572 and activity levels from vowel segments of continuous speech. *Journal of Voice* **24**: 30-38.

573 **Wood R, Gustafson G. 2001.** Infant Crying and Adults' Anticipated Caregiving Responses:

574 Acoustic and Contextual Influences. *Child Development* **72(5)**: 1287-1300.

Pair comparison	<b>t</b>	<b>P</b>
<i>Acoustic Dimension AD1</i>		
Chimp/Bonobo	-3.10	0.017
Human (bath)/Bonobo	-10.18	<0.001
Human (vaccine)/Bonobo	-6.82	<0.001
Human (bath)/Chimp	-7.08	<0.001
Human (vaccine)/Chimp	-3.72	0.003
Human (vaccine)/Human (bath)	3.37	0.008
<i>Acoustic Dimension AD2</i>		
Chimp/Bonobo	-2.96	0.025
Human (bath)/Bonobo	-0.168	0.998
Human (vaccine)/Bonobo	2.63	0.054
Human (bath)/Chimp	2.79	0.038
Human (vaccine)/Chimp	5.59	<0.001
Human (vaccine)/Human (bath)	2.80	0.037

**Table 1. Results of the multiple comparison tests for the Acoustic Dimensions AD1 and AD2.** The table presents raw P values; the pair comparisons that are significant after Bonferroni correction are shown in bold.

<b>Acoustic parameters</b>	<b>Acoustic Dimension AD1 (% of variance = 37%)</b>	<b>Acoustic Dimension AD2 (% of variance = 21%)</b>
%voiced	0.441	-0.353
mean pitch	<b>0.897</b>	-0.120
max pitch	<b>0.780</b>	0.016
min pitch	0.450	-0.384
pitchCV	-0.305	0.424
inflex25	0.579	-0.113
inflex2	0.051	-0.182
intCV	<b>-0.779</b>	0.417
harm	<b>-0.787</b>	0.045
jitter	<b>0.845</b>	-0.030
shimmer	<b>0.856</b>	-0.261
SP1	0.398	<b>0.677</b>
SP2	0.410	<b>0.825</b>
SP3	0.429	<b>0.829</b>
SP4	0.342	<b>0.800</b>

**Table 2. Factor loadings on the two acoustic dimensions calculated from the acoustic parameters describing infants' calls in bonobos, chimpanzees and humans.** The first principal component is mainly related to both the pitch and the periodic quality of the signal. The second principal component is mainly related to the distribution of the energy among the frequency spectrum.

	voiced (%)	meanpitch (Hz)	maxpitch (Hz)	minpitch (Hz)	pitchCV	inflex25	Inflex2	intCV	harm	jitter	shimmer	SP1 (Hz)	SP2 (Hz)	SP3 (Hz)	SP4 (Hz)
Human (bath)	43.6 ± 10.1	607 ± 211	2201 ± 716	246 ± 93	0.484 ± 0.157	14.4 ± 2.9	1.87 ± 0.69	16.8 ± 0.97	14.26 ± 3.42	0.010 ± 0.008	0.070 ± 0.038	1524 ± 480	3094 ± 804	5117 ± 1420	7689 ± 1380
Human (vaccine)	26.4 ± 11.6	1132 ± 600	2835 ± 823	228 ± 122	0.638 ± 0.196	14.3 ± 4.9	1.46 ± 0.74	16.2 ± 1.25	9.19 ± 1.48	0.023 ± 0.011	0.111 ± 0.045	1740 ± 615	4307 ± 1985	7133 ± 2273	10811 ± 2674
Bonobo	56.1 ± 10.6	2681 ± 445	3750 ± 242	460 ± 293	0.363 ± 0.199	19.5 ± 3.7	1.97 ± 0.65	0.970 ± 0.262	6.56 ± 2.32	0.032 ± 0.010	0.169 ± 0.026	1924 ± 1164	4438 ± 1775	7339 ± 1833	10705 ± 3232
Chimpanzee	49.7 ± 18.4	1588 ± 262	3007 ± 724	648 ± 329	0.275 ± 0.162	18.1 ± 4.9	1.75 ± 0.74	0.894 ± 0.187	7.82 ± 2.19	0.027 ± 0.009	0.171 ± 0.038	1523 ± 257	2873 ± 420	5011 ± 1574	7432 ± 2871

**Table 3. Acoustic characteristics of calls.** The table reports the means ( $\pm$  s.d.) of measured variables.



Acoustic parameters	Estimate	s.e.	t	$\chi^2$	P
Intercept	4.000	0.196	20.2		
<b>AD1</b>	0.298	0.059	5.05	24.39	<b>&lt; 0.001</b>
AD2	0.075	0.076	0.99	4.23	0.375
Listener's gender	0.164	0.165	1.00	3.37	0.498
AD1*AD2	-0.038	0.038	-1.00	1.74	0.419
AD1*gender	-0.001	0.043	-0.03	1.72	0.632
AD2*gender	0.039	0.044	0.88	3.37	0.339
AD1*AD2*gender	0.021	0.020	1.02	3.39	0.495
Intercept	4.024	0.165	24.4		
%voiced	-0.03	0.178	-0.20	0.00	1.000
<b>mean pitch</b>	0.950	0.282	3.37	13.90	<b>&lt; 0.001</b>
max pitch	-0.445	0.247	-1.80	4.10	0.043
min pitch	0.080	0.159	0.50	0.00	1.000
pitchCV	0.329	0.200	1.65	3.41	0.065
inflex25	-0.045	0.149	-0.31	0.00	1.000
inflex2	0.120	0.149	0.80	0.50	0.478
intCV	-0.388	0.240	-1.62	3.28	0.070
harm	-0.263	0.306	-0.86	0.65	0.422
jitter	0.477	0.327	1.46	2.59	0.108
shimmer	-0.778	0.330	-2.36	7.12	0.008
SP1	-0.108	0.229	-0.47	0.00	1.000
SP2	-0.286	0.361	-0.79	0.48	0.489
SP3	0.338	0.392	0.86	0.64	0.424
SP4	0.184	0.271	0.68	0.25	0.617

**Table 4. Acoustic predictors of listeners' rating.** Two separate models were built: the first model had the two acoustic dimensions defined by the Principal Component Analysis (AD1 and AD2) as fixed variables; the second model had the whole set of 15 temporal and spectral parameters (z-transformed) as fixed variables. Listeners' rating was the dependent measure in both models. P values were obtained with likelihood-ratio tests comparing the fit of full models with reduced models lacking the fixed effect. The table presents raw P values; the significant predictors after Bonferroni correction are shown in bold.

## FIGURES LEGEND

**Figure 1. Spectrograms of acoustic signals.** Top left: Human baby cry recorded in the “bath” context. Bottom left: Human baby cry recorded in the “vaccine” context. Top right: Chimpanzee infant call. Bottom right: Bonobo infant call. Graphical representation produced with the R package Seewave (Sueur et al. 2008) with spectrograms set to Hanning window and a FFT window length of 1024 with 70% overlap. See Supplementary Recordings to hear the sounds.

**Figure 2. Distribution of distress calls within the acoustic space.** Each dot represents a single recording. The acoustic structure of calls was first described using 15 parameters in the frequency and temporal domains, and further reduced using a principal component analysis into two independent Acoustic Dimensions (AD1 and AD2). The first axis of the acoustic space (AD1) is mainly related to both the pitch and the periodic quality of the signal. The second axis of the acoustic space (AD2) is mainly related to the distribution of the energy among the frequency spectrum.

**Figure 3. Cry pitch drives the perceived distress by adult listeners in human and ape infants’ cries.** Blue dots = human baby cries (“bath” and “vaccine” cries are not distinguished here); yellow dots = chimpanzee infant calls; red = bonobo infant calls. Solid curves represent fits ( $y = \log(\text{mean pitch})$ ) of the estimated marginal means  $\pm$  SE (black curve: all three species confounded).

## **SUPPORTING INFORMATION**

**Supplementary Recording S1 (Human\_bath.wav):** Sound recording of a human baby's cry, recorded in the bathing context.

**Supplementary Recording S2 (Human\_vaccine.wav):** Sound recording of a human baby's cry, recorded in the vaccination context.

**Supplementary Recording S3 (Bonobo.wav):** Sound recording of a bonobo baby's call

**Supplementary Recording S4 (Chimp.wav):** Sound recording of a chimpanzee baby's call